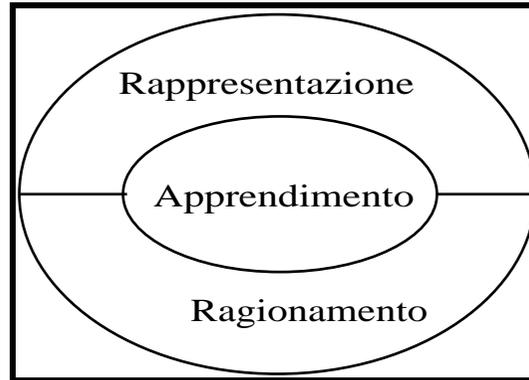


# APPRENDIMENTO

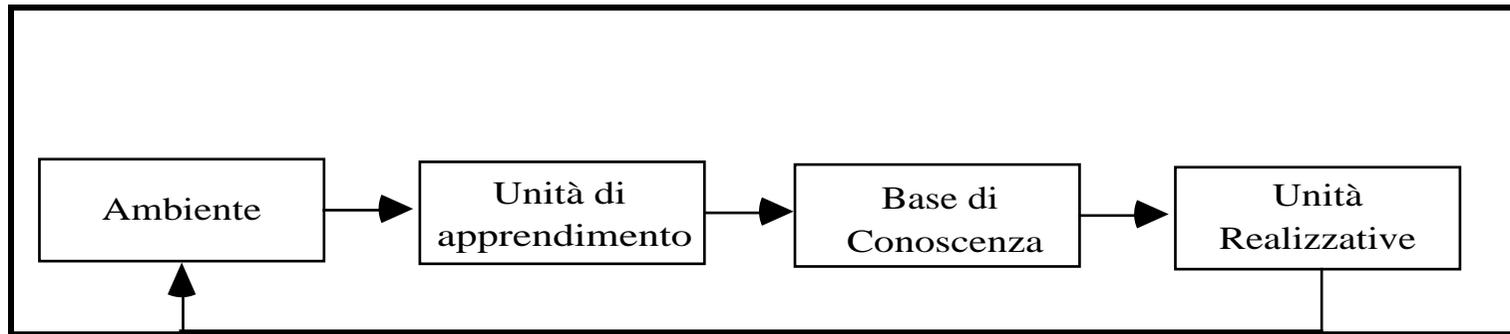
- **UNO DEGLI ELEMENTI COSTITUENTI L'INTELLIGENZA**



- **DUE TIPI DI APPRENDIMENTO**
  - **DEDUTTIVO:** LE REGOLE GENERALI VENGONO USATE PER RICAVARE FATTI SPECIFICI (ES. THEOREM PROVING)
  - **INDUTTIVO:** LE REGOLE GENERALI VENGONO RICAVATE DALL'OSSERVAZIONE DI FATTI SINGOLI (ES. RICONOSCIMENTO)

# APPRENDIMENTO INDUTTIVO

- **PROCESSO DI ACQUISIZIONE DI NUOVA CONOSCENZA TRAMITE ESPERIENZA**



- **APPRENDIMENTO COME OTTIMIZZAZIONE DI UNA FUNZIONE OBIETTIVO**

## **APPRENDIMENTO INDUTTIVO (CONT.)**

- **DUE CLASSI DI ARCHITETTURE:**
  - **SISTEMI SIMBOLICI:**
    - ENFASI SU RAPPRESENTAZIONE E RAGIONAMENTO DI TIPO LOGICO
    - LE COMPONENTI SONO STRUTTURE DI ALTO LIVELLO (PROPOSIZIONI, REGOLE,...)
    - MODELLO DI COMPUTAZIONE SEQUENZIALE
  - **SISTEMI SUB-SIMBOLICI O CONNESSIONISTICI**
    - ENFASI SU RAPPRESENTAZIONE DI TIPO NUMERICO
    - LE COMPONENTI SONO SEMPLICI E NUMEROSE
    - MODELLO DI COMPUTAZIONE PARALLELO
- **DUE TIPI DI APPRENDIMENTO**
  - **APPRENDIMENTO SUPERVISIONATO (CLASSIFICAZIONE, LEARNING BY EXAMPLE):**
    - VIENE DEFINITO L'INSIEME DI CLASSI DEL PROBLEMA E UN INSIEME DI ESEMPI PRE-CLASSIFICATI
    - IL SISTEMA DEVE TROVARE UNA CARATTERIZZAZIONE DELLE CLASSI CHE PERMETTA LA CORRETTA CLASSIFICAZIONE DI ESEMPI NON VISTI (GENERALIZZAZIONE)
  - **APPRENDIMENTO NON SUPERVISIONATO (CLUSTERING, LEARNING BY OBSERVATION):**
    - GLI OGGETTI ESEMPIO SONO FORNITI PRIVI DI UNA CLASSIFICAZIONE A PRIORI
    - IL SISTEMA DEVE TROVARE DELLE REGOLARITÀ O SOMIGLIANZE

## **ESEMPI DI SISTEMI DI APPRENDIMENTO**

- **RETI NEURALI**
- **ALGORITMI GENETICI**
- **SISTEMI A REGOLE**
- **ALBERI DI DECISIONE**
- **RETI BAYESIANE**

# DIFFERENZE TRA DATA MINING E MACHINE LEARNING

- SOVRAPPOSIZIONE NEL TIPO DI PROBLEMI CONSIDERATI E SOLUZIONI ADOTTATE
- DIFFERENZE PRINCIPALI
  - ML ORIENTATA AL MIGLIORAMENTO DELLE PRESTAZIONI, DM ORIENTATA ALLA PRODUZIONE DI CONOSCENZA *COMPENSIBILE*
  - QUESTIONI DI EFFICIENZA PIÙ RILEVANTI PER DM CHE PER ML, LEGATE ALLA DIMENSIONE DELL'INSIEME DI DATI
  - DM SI OCCUPA DI DATI E UTILIZZA DATABASE REALI
    - INFORMAZIONI ASSENTI/PARZIALI/INCONSISTENTI
    - ERRORI
  - ML COMPRENDE PROBLEMATICHE PIÙ AMPIE

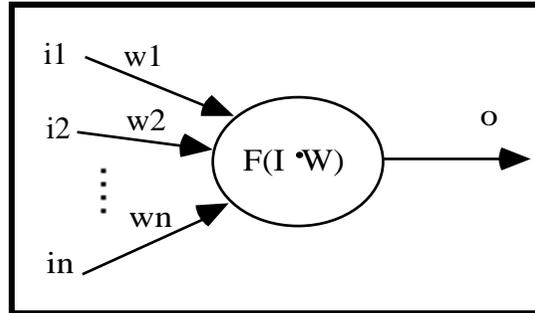
# **PROBLEMI E TEMATICHE DI INTERESSE**

## **PROBLEMI LEGATI ALLA PRESENZA DI UN DATABASE REALE:**

- **INFORMAZIONE LIMITATA**
  - **ASSENZA DI ATTRIBUTI NECESSARI ALL'APPRENDIMENTO**
- **RUMORE E DATI MANCANTI**
  - **ERRORI DI INSERIMENTO, CAMPI NULLI, ...**
  - **VARIE TECNICHE DI GESTIONE DEL RUMORE**
  - **TRATTAMENTO DELL'INCERTEZZA**
- **DIMENSIONI**
  - **MULTI-GIGABYTES, TERABYTES SONO ORMAI REALTÀ**
  - **SCALABILITÀ DEGLI ALGORITMI, PARALLELISMO E DISTRIBUZIONE**
- **DINAMICITÀ**
  - **COME MANTENERE AGGIORNATA E CONSISTENTE L'INFORMAZIONE**
  - **GESTIONE DEL TEMPO**
- **CAMPI IRRILEVANTI**
  - **FUNZIONE DELLO SCOPO DEL DM**
  - **FEATURE SELECTION E FEATURE REDUCTION**
- **ETEROGENITÀ DEI DATI**
  - **SE I DATI SONO CONTENUTI IN DATABASE DIVERSI**
  - **SINONIMIE, OMONIMIE, SCALA, .....**
- **PROBLEMI DI PRIVACY**

# RETI NEURALI

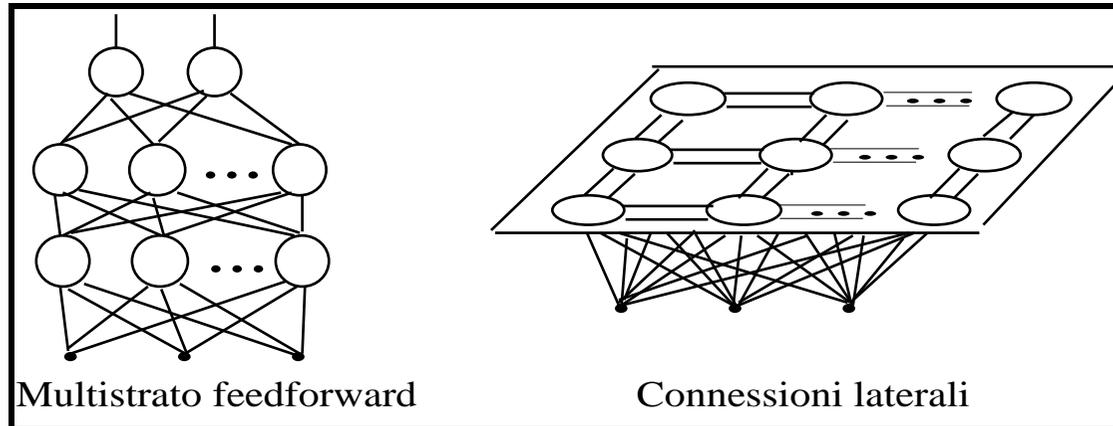
- Insieme di semplici elementi di calcolo (neuroni)



$$I \cdot W = i_1 \cdot w_1 + i_2 \cdot w_2 + \dots + i_n \cdot w_n$$

$F(\cdot)$  = funzione non lineare (es. gradino, sigmoide)

- Connessi tra loro secondo qualche topologia di rete: l'output di un neurone diventa l'input (pesato) per altri neuroni
- Diverse topologie di rete



Multistrato feedforward

Connessioni laterali

# APPRENDIMENTO NELLE RETI NEURALI

**Processo mediante il quale i parametri liberi di una rete neurale vengono modificati.**

**parametri liberi = pesi delle connessioni tra neuroni**

**Le modificazioni avvengono mediante un processo continuo di stimolazione da parte dell'ambiente nel quale la rete è immersa.**

$$w_{kj}^{(t+1)} = F(x^{(t)}, W^{(t)})$$

**Il tipo di apprendimento è determinato dal modo in cui i parametri vengono modificati**

- **Hebbiano**
- **Competitivo**
- **Boltzman**
- **Back Prop**

**Questo a sua volta dipende dalla funzione da ottimizzare**

# NATURA STATISTICA DEL PROCESSO DI APPRENDIMENTO

## APPRENDIMENTO

=

PROBLEMA DI COME TROVARE UNA DIPENDENZA FUNZIONALE USANDO UN NUMERO LIMITATO DI OSSERVAZIONI

- $X$  variabile aleatoria  $p$ -dimensionale con p.d.f.  $P_X(x)$ ,
- $x_1, x_2, \dots, x_m \in \mathcal{X}^p$ ,  $m$  realizzazioni di  $X$
- $Y = g(X)$ ,  $Y \in \mathcal{Y}$ ,  $g(\cdot)$  non è nota a priori
  - $g(\cdot)$  funzione non deterministica:  $P_{Y/X}(y/x)$  generica

- $g(\cdot)$  funzione deterministica:  $P_{Y/X}(y/x) = \begin{cases} 1 & \text{se } y = g(x) \\ 0 & \text{altrimenti} \end{cases}$

- $y_1, y_2, \dots, y_m$ ,  $y_k = g(x_k)$

TROVARE LA MIGLIORE APPROSSIMAZIONE DELLA FUNZIONE  $g(\cdot)$

=

TROVARE LA FUNZIONE  $f(x, \theta)$  CHE MINIMIZZA IL FUNZIONALE DI PERDITA

$$R(\theta) = \int L(y, f(x, \theta)) dP(x, y) = \int L(y, f(x, \theta)) p(x, y) dx dy$$

- $L(y, f(x, \theta))$  funzione che misura la perdita o discrepanza tra la funzione effettiva e la sua approssimazione

## STIMA DELLA FUNZIONE DI REGRESSIONE

- **modello regressivo:**  $Y = g(X) + \varepsilon$ ,
- **Soluzione nel modello regressivo:**  $\hat{g}(X) = E[Y|X] = \int y dP(y | x)$ ,
- $\hat{g}(X)$  è la funzione che minimizza la *distorsione quadratica media*

↓

$$\hat{g}(X) = \text{MIN } R(\Theta), \text{ in cui } L(y, f(x, \Theta)) = (y - f(x, \Theta))^2$$

- **APPLICAZIONI TIPICHE NELL'AMBITO DEL TRATTAMENTO DI SEGNALI**
  - **COMPRESSIONE**
  - **TRASMISSIONE SU CANALE ADDITIVO**

## RICONOSCIMENTO DI PATTERN

- **modello :**  $p(x) = \sum_{i=1}^k p(x | c_i)P(c_i)$ ,  $\{c_1, c_2, \dots, c_k\}$  insieme di etichette  
le classi non sono separate:  $p(c_i/x) \neq 1$

trovare un classificatore (o regola di decisione)  $g : \mathcal{R}^P \rightarrow \{c_1, c_2, \dots, c_k\}$  che minimizzi la probabilità di errore  $R(g) = \int R(g(x) | x) dP(x)$ , con

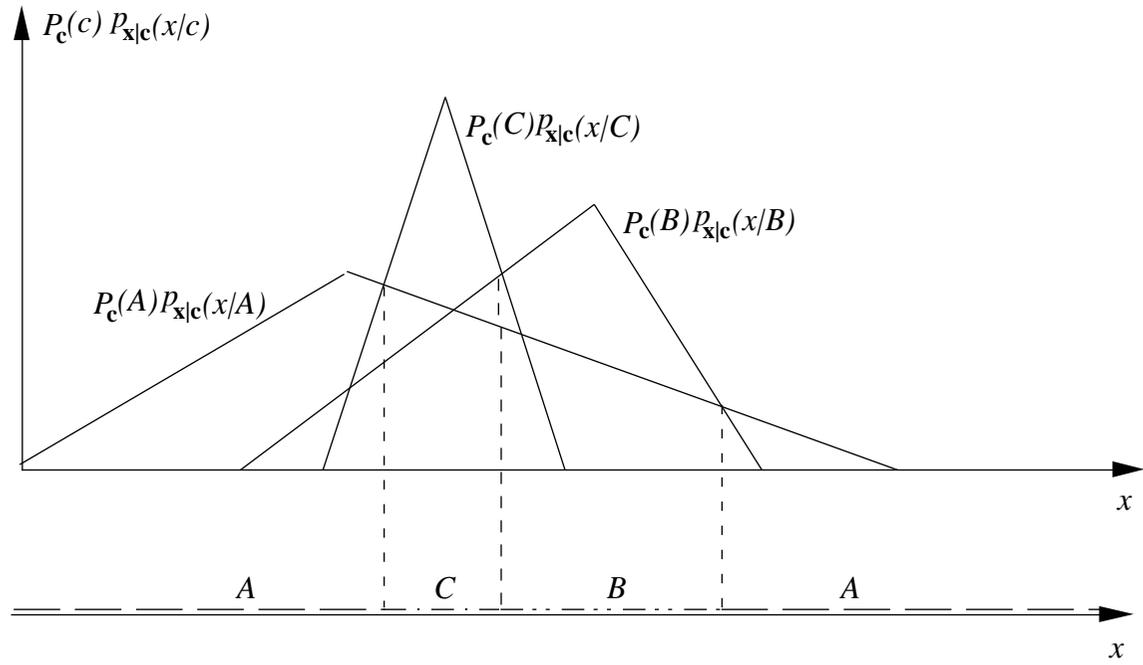
$$R(c_j | x) = \sum_{i \neq j} P(c_i | x),$$

- **soluzione:** regola di decisione Bayesiana  $\hat{g}(x) = \text{MIN}_{j=1, \dots, k} R(c_j/x)$ ,
- in questo caso

$$\begin{aligned} R(\theta) &= \int \sum_{c_i \neq g(x)} p(c_i | x) dP(x), = \int \sum_{c_i \neq g(x)} p(c_i | x) p(x) dx \\ &= \int \sum_{c_i \neq g(x)} p(c_i, x) dx = \sum_{i=1}^k \int L(y, g(x)) p(c_i, x) dx \end{aligned}$$

da cui si ricava

$$L(y, g(x, \theta)) = \begin{cases} 1 & \text{se } y = g(x) \\ 0 & \text{altrimenti} \end{cases}$$



# APPRENDIMENTO

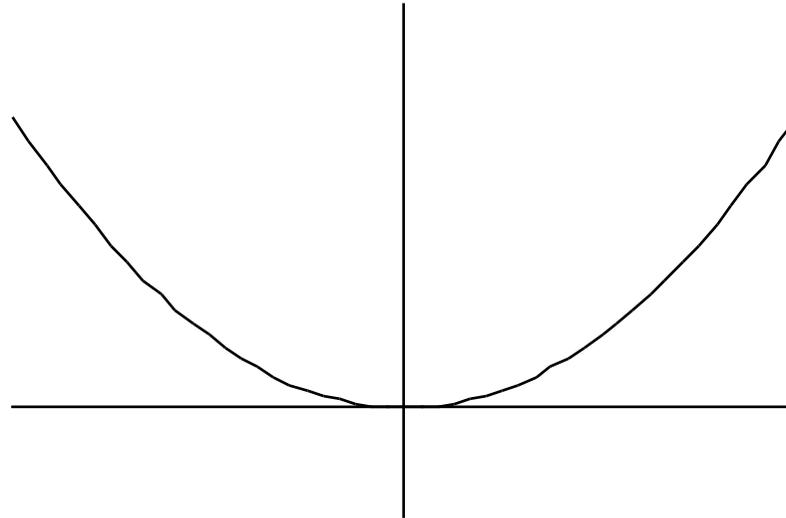
=

## MINIMIZZAZIONE DI UNA FUNZIONE OBIETTIVO

- ALGORITMI DI OTTIMIZZAZIONE  $\Rightarrow$ 

ALGORITMI DI GRADIENTE (OTTIMI LOCALI)
ALGORITMI DI RILASSAMENTO STOCASTICO (OTTIMO GLOBALE)
- ALGORITMO DI GRADIENTE (DISCENDENTE):  $\Theta^{(k+1)} = \Theta^{(k)} - \gamma \nabla_{\Theta} F(\Theta^{(k)})$ 
  - $F(\cdot)$ : FUNZIONE DA OTTIMIZZARE
  - $\Theta$ : PARAMETRO DAL QUALE DIPENDE F
  - $\nabla_{\Theta}$ : VETTORE DELLE DERIVATE PARZIALI RISPETTO A  $\Theta$

$$F(x)=x^2$$



- **PROBLEMI CON L'ALGORITMO DI GRADIENTE:**

- **OTTIMIZZAZIONE UNIFORME PER TUTTE LE COORDINATE  $\Rightarrow$  GRADIENTE PER COORDINATE**

<ul style="list-style-type: none"><li>• <b>CALCOLO DEL GRADIENTE AD OGNI ITERAZIONE COSTOSO</b></li><li>• <b>FUNZIONE DA MINIMIZZARE NON NOTA</b></li></ul>	<b>} <math>\Rightarrow</math> GRADIENTE STOCASTICO</b>
---	--

# METODI DI APPROSSIMAZIONE STOCASTICA

- **PROBLEMA:** MINIMIZZARE IL VALORE DI UNA FUNZIONE DETERMINISTICA  $F(\Theta)$  DI UN PARAMETRO  $\Theta$  NON NOTA
- **IPOTESI:** È POSSIBILE OSSERVARE, PER OGNI  $\Theta$ , IL VALORE DI UNA VARIABILE ALEATORIA  $Y(\Theta)$  TALE CHE  
 $E[Y(\Theta)] = \nabla_{\Theta} F(\Theta)$  (STIMATORE CORRETTO O NON DISTORTO)
- ALLORA UN ALGORITMO ITERATIVO DELLA FORMA

$$\Theta^{(k+1)} = \Theta^{(k)} - \gamma^{(k)} Y(\Theta^{(k)}) \quad (\text{GRADIENTE STOCASTICO})$$

CONVERGE AD UN VALORE  $\Theta^*$  CHE È UN MINIMO (LOCALE) DELLA FUNZIONE  $F(\Theta)$

- CONVERGENZA ASINTOTICA

$$\lim_{k \rightarrow \infty} E[(\theta^{(k)} - \theta^*)^2] = 0$$

$$\lim_{k \rightarrow \infty} P(\theta^{(k)} = \theta^*) = 1$$

- CONDIZIONI SU  $\gamma^{(k)}$
- IN GENERALE LA STIMA  $\hat{Y}$  È BASATA SU UN INSIEME DI OSSERVAZIONI SPERIMENTALI